On the Power of Cooperation: Can a Little Help a Lot? (Extended Version)

Parham Noorzad*, Michelle Effros*, Michael Langberg†, Tracey Ho*
*California Institute of Technology
Emails: {parham, effros, tho}@caltech.edu
†State University of New York at Buffalo
Email: mikel@buffalo.edu

***Abstract*—In this paper, we propose a new cooperation model for discrete memoryless multiple access channels. Unlike in prior cooperation models (e.g., conferencing encoders), where the transmitters cooperate directly, in this model the transmitters cooperate through a larger network. We show that under this indirect cooperation model, there exist channels for which the increase in sum-capacity resulting from cooperation is significantly larger than the rate shared by the transmitters to establish the cooperation. This result contrasts both with results on the benefit of cooperation under prior models and results in the network coding literature, where attempts to find examples in which similar small network modifications yield large capacity benefits have to date been unsuccessful.**

I. Introduction

Cooperation is a potentially powerful strategy in distributed communication systems. It can both increase the possible transmission rates of source messages and improve the reliability of network communications [1]. To date, cooperation is not completely understood. In this paper, we focus on the effect of cooperation on the capacity region and discuss situations where a small amount of rate used to enable cooperation results in a large increase in the total information that can be carried by the network.

One model of cooperation, proposed by Willems in [2], is the *conferencing encoders* (CE) model for the discrete memoryless multiple access channel (DM-MAC). In the CE model, there is a noiseless link of capacity C_{12} from the first encoder to the second and a corresponding link of capacity C_{21} back. These links allow a finite number of rounds of communication between the two encoders; the total number of bits sent by each encoder to the other is bounded by the product of the DM-MAC coding blocklength and the capacity of the encoder's outgoing cooperation link. A similar type of cooperation is applied in the broadcast channel with conferencing decoders [3] and the interference channel with conferencing encoders [4]. More recently, the authors of [5] investigate the case where each encoder has partial state information and conferencing enables information exchange about both the state and the messages.

One can imagine scenarios in which the two transmitters are not able to communicate directly or can communicate more effectively through some other part of the network. The latter can occur, for example, if resources are less constrained elsewhere in the network than they are for direct communication. To capture such scenarios, we introduce the *cooperation facilitator* (CF) model for the DM-MAC. The cooperation facilitator is a node that has complete access to both source messages. Based on the messages, it sends limited-rate information to both encoders through a noiseless bottleneck link of finite capacity (Figure 1). We define the *cooperation rate* as the capacity of the link carrying the information to be shared. One can think of capacity gains obtained from this model as an outer bound on the benefit of indirect cooperation.

To study cooperation under this model, we compare the sum-capacity of a DM-MAC with a CF to the sum-capacity of the DM-MAC when there is no cooperation between the transmitters. This difference equals the capacity cost of removing the CF output link from the network. When the link is removed, the two transmitters are not able to cooperate, and their transmitted codewords are independent. We call the resulting network the DM-MAC with *independent encoders* (IE). The capacity region of this network is due to Ahlswede [6], [7] and Liao [8].

Since removing the bottleneck link transforms the CF network into the IE network, the proposed cooperation model is related to the edge removal problem in network coding [9]–[13]. For networks of noiseless links, there are no known examples of networks for which removing a single edge of capacity δ changes the capacity region by more than δ in each dimension, and in some cases it is known that an impact of more than δ per dimension is not possible [9], [10]. Therefore, at least in the situations investigated in [9], [10], inserting a cooperation facilitator in a network cannot increase the sum-capacity by more than a constant times the cooperation rate.

How much can cooperation help in a DM-MAC? In the CE model, the increase in sum-capacity is at most the sum of the capacities of the noiseless links between the two encoders (Section II). Given the previous discussion, one may wonder whether a similar result holds for the CF model, that is, whether the increase in sum-capacity is limited to a constant times the cooperation rate. In what follows, we see that the benefit of cooperation can far exceed what might be expected based on the CE and edge removal examples. Specifically, we describe a sequence of DM-MACs with increasing alphabet sizes and set the cooperation rate for each channel as a function of its alphabet size. We then show that the increase in sum-capacity that results from cooperation grows more quickly than any polynomial function of the cooperation rate.

In the next section, we review the CE model and its capacity region as presented by Willems [2]. We give a formal introduction to the CF model in Section III.

II. Prior Work

Consider the DM-MAC

$$\left(\mathcal{X}_1 \times \mathcal{X}_2, p_{Y|X_1,X_2}(y|x_1,x_2), \mathcal{Y}\right),$$

where $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ are finite sets and $p_{Y|X_1,X_2}(y|x_1,x_2)$ denotes the conditional distribution of the output, Y, given the inputs, X_1 and X_2. To simplify notation, we suppress the subscript of the probability distributions when the corresponding random variables are clear from context. For example, we write $p(x)$ instead of $p_X(x)$.

There are two sources, source 1 and source 2, whose outputs are the messages $W_1 \in \mathcal{W}_1 = \left\{1, \ldots, \left\lceil 2^{nR_1} \right\rceil\right\}$ and $W_2 \in \mathcal{W}_2 = \left\{1, \ldots, \left\lceil 2^{nR_2} \right\rceil\right\}$, respectively. The random variables W_1 and W_2 are independent and uniformly distributed over their corresponding alphabets. The real numbers R_1 and R_2 are nonnegative and are called the *message rates*.

In the IE model, each encoder only has access to its corresponding message. The encoders are represented by the functions

$$f_{1n} : \mathcal{W}_1 \to \mathcal{X}_1^n,$$
$$f_{2n} : \mathcal{W}_2 \to \mathcal{X}_2^n.$$

We denote the output of the encoders by $X_1^n = f_{1n}(W_1)$ and $X_2^n = f_{2n}(W_2)$. Let Y^n be the output of the channel when the pair (X_1^n, X_2^n) is transmitted. Using Y^n, the decoder estimates the original messages via a decoding function $g_n : \mathcal{Y}^n \to \mathcal{W}_1 \times \mathcal{W}_2$.

A $\left(2^{nR_1}, 2^{nR_2}, n\right)$ code for the multiple access channel is defined as the triple (f_{1n}, f_{2n}, g_n). The average probability of error for this code is given by

$$P_e^{(n)} = \Pr\left(g_n\left(Y^n\right) \neq (W_1, W_2)\right).$$

We say the rate pair (R_1, R_2) is achievable if there exists a sequence of $\left(2^{nR_1}, 2^{nR_2}, n\right)$ codes such that $P_e^{(n)}$ tends to zero as the blocklength, n, approaches infinity. The capacity region, $\mathscr{C}$, is the closure of the set of all achievable rate pairs.

For a given capacity region $\mathscr{C} \subseteq \mathbb{R}^2_{\geq 0}$, the *sum-capacity* [14], C_S, is defined as

$$C_\mathrm{S} = \max\left\{R_1 + R_2 \middle| (R_1, R_2) \in \mathscr{C}\right\}. \tag{1}$$

In the IE model [6]–[8], the sum-capacity is given by

$$C_{\mathrm{S-IE}} = \max_{p(x_1)p(x_2)} I\left(X_1, X_2; Y\right).$$

In the CE model, each encoder shares some information regarding its message with the other encoder prior to transmission over the channel. This sharing of information is achieved through a *K-step conference* over noiseless links of capacities C_{12} and C_{21}. A K-step conference consists of two sets of functions, $\{h_{11}, \ldots, h_{1K}\}$ and $\{h_{21}, \ldots, h_{2K}\}$, which recursively define the random vectors $V_1^K := (V_{11}, \ldots, V_{1K})$ and $V_2^K := (V_{21}, \ldots, V_{2K})$ as

$$V_{1k} = h_{1k}\left(W_1, V_2^{k-1}\right),$$
$$V_{2k} = h_{2k}\left(W_2, V_1^{k-1}\right)$$

for $k = 1, \ldots, K$. In step k, encoder 1 (encoder 2) computes V_{1k} (V_{2k}) and sends it to encoder 2 (encoder 1). Since the noiseless links between the two encoders are of capacity C_{12} and C_{21}, respectively, we require

$$\sum_{k=1}^{K} \log|\mathcal{V}_{1k}| \leq nC_{12},$$
$$\sum_{k=1}^{K} \log|\mathcal{V}_{2k}| \leq nC_{21}$$

where $\mathcal{V}_{ik}$ is the alphabet of the random variable V_{ik} for $i = 1, 2$ and $k = 1, \ldots, K$. The outputs of the encoders, X_1^n and X_2^n, are given by

$$X_1^n = f_{1n}\left(W_1, V_2^K\right),$$
$$X_2^n = f_{2n}\left(W_2, V_1^K\right)$$

where f_{1n} and f_{2n} are deterministic functions.

By studying the capacity region of the CE model [2], we deduce

$$C_{\mathrm{S-IE}} \leq C_{\mathrm{S-CE}} \leq C_{\mathrm{S-IE}} + C_{12} + C_{21}.$$

Thus, with conferencing, the sum-capacity increases at most linearly in (C_{12}, C_{21}) over the sum-capacity of the IE model.

III. The Cooperation Facilitator: Model and Result

In the CF model, cooperation is made possible not through finite capacity links between the encoders, but instead through a third party, the cooperation facilitator, which receives information from both encoders and transmits a single description of that information back to both (Figure 1). The cooperation facilitator is represented by the function

$$\phi_n : \mathcal{W}_1 \times \mathcal{W}_2 \to \mathcal{Z},$$

where the alphabet $\mathcal{Z} = \left\{1, \ldots, \left\lceil 2^{n\delta} \right\rceil\right\}$ is determined by the cooperation rate δ. The output of the cooperation facilitator, $Z = \phi_n(W_1, W_2)$, is available to both encoders. Each encoder chooses a blocklength-n codeword as a function of its own source and Z and sends that codeword to the receiver using n transmissions. Hence the two encoders are represented by the functions

$$f_{1n} : \mathcal{W}_1 \times \mathcal{Z} \to \mathcal{X}_1^n,$$
$$f_{2n} : \mathcal{W}_2 \times \mathcal{Z} \to \mathcal{X}_2^n.$$

The definitions of the decoder, probability of error, and capacity region are similar to the IE model discussed in the previous section and are omitted.

Given a pair of functions $f, g : \mathbb{Z}^+ \to \mathbb{R}$, we say $f = o(g)$ if $\lim_{m\to\infty} \frac{f(m)}{g(m)} = 0$. We say $f = \omega(g)$ if $g = o(f)$.

For a sequence of DM-MACs

$$\left\{\left(\mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)}, p^{(m)}(y|x_1,x_2), \mathcal{Y}^{(m)}\right)\right\}_m,$$

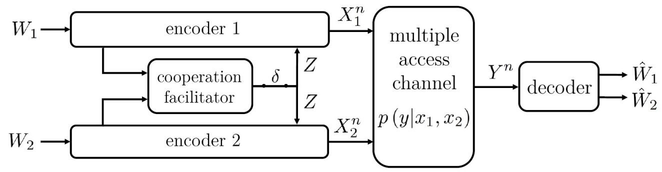

Figure 1. Network model for the DM-MAC with a CF. The *cooperation rate* is the capacity of the output link of the CF which we denote by δ.

let $C^{(m)}_{\mathrm{S-IE}}$ denote the IE sum-capacity of the m^{th} channel and $C^{(m)}_{\mathrm{S-CF}}$ denote the CF sum-capacity of the m^{th} channel when the cooperation rate is δ_m.

We are now ready to answer the question posed in the introduction. In the next theorem, which is the main result of this paper, we see that for a sequence of DM-MACs, the increase in sum-capacity is not only greater than the cooperation rate, but also asymptotically larger than any polynomial function of that rate. In what follows, $\log(x)$ is the base 2 logarithm of x.

Theorem 1. *For every sequence of cooperation rates $\{\delta_m\}_m$ satisfying $\delta_m = \log m + \omega(1)$ and $\delta_m \le m$ and every $\epsilon > 0$, there exists a sequence of DM-MACs with input alphabets*

$$\mathcal{X}_1^{(m)} = \mathcal{X}_2^{(m)} = \{1, \ldots, 2^m\},$$

such that for sufficiently large m,

$$C^{(m)}_{\mathrm{S-CF}} - C^{(m)}_{\mathrm{S-IE}} \ge (3 - \sqrt{5+4\epsilon})m - \delta_m.$$

For the same sequence of channels, we also have

$$C^{(m)}_{\mathrm{S-CF}} - C^{(m)}_{\mathrm{S-IE}} \le m + \delta_m.$$

In the above theorem, the choice of δ_m is constrained only by $\delta_m = \log m + \omega(1)$ and $\delta_m \le m$. For example, a cooperation rate of $\delta_m = \log(m \log m)$ can lead to an increase in sum-capacity that is linear in m, giving a capacity benefit that is "almost" exponential in the cooperation rate.

In the next section, we prove the existence of a sequence of DM-MACs with properties that are essential for the proof of Theorem 1. In Section V, we show that for the sequence of channels of Section IV,

$$2m - \delta_m \le C^{(m)}_{\mathrm{S-CF}} \le 2m. \tag{2}$$

In Section VI we show

$$m - \delta_m \le C^{(m)}_{\mathrm{S-IE}} \le (\sqrt{5+4\epsilon} - 1)m. \tag{3}$$

Combining these two results gives Theorem 1. See Figure 2.

IV. Channel Construction

For a fixed positive integer m, the channel

$$\left(\mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)}, p^{(m)}(y|x_1,x_2), \mathcal{Y}^{(m)}\right)$$

used in the proof of Theorem 1 has input alphabets $\mathcal{X}_1^{(m)} = \mathcal{X}_2^{(m)} = \{1, \ldots, 2^m\}$ and output alphabet

$$\mathcal{Y}^{(m)} = \left(\mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)}\right) \cup \{(E,E)\},$$

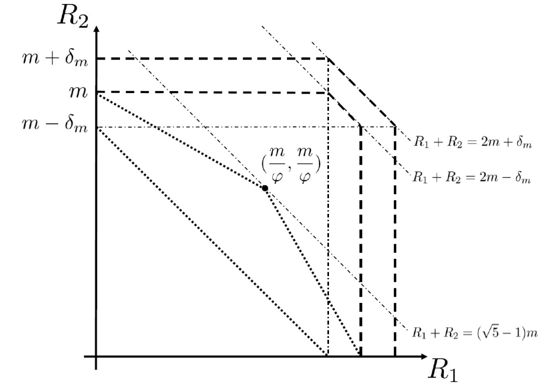

Figure 2. Inner and outer bounds for the capacity regions of the CF (dashes) and IE (dots) models as derived in Sections V and VI, respectively. In this figure, $\varphi = \frac{1+\sqrt{5}}{2}$.

where "E" denotes an erasure symbol. For each $(x_1, x_2, y) \in \mathcal{X}_1^{(m)} \times \mathcal{X}_2^{(m)} \times \mathcal{Y}^{(m)}$, $p^{(m)}(y|x_1,x_2)$ is defined as a function of the corresponding entry $b_{x_1x_2}$ of a binary matrix $B_m = (b_{ij})_{i,j=1}^{2^m}$. Precisely,

$$p^{(m)}(y|x_1,x_2) = \begin{cases} 1 - b_{x_1x_2}, & \text{if } y = (x_1,x_2) \\ b_{x_1x_2}, & \text{if } y = (E,E). \end{cases} \tag{4}$$

That is, when (x_1,x_2) is transmitted, $y = (x_1,x_2)$ is received if $b_{x_1x_2} = 0$, and $y = (E,E)$ is received if $b_{x_1x_2} = 1$. Thus, we interpret the 0 and 1 entries of B_m as "good" and "bad" entries, respectively. Let $\mathcal{X}^{(m)} = \{1, \ldots, 2^m\}$. We define the sets

$$0_{B_m} = \{(i,j) : b_{ij} = 0\},$$
$$1_{B_m} = \{(i,j) : b_{ij} = 1\}$$

to be the set of good and bad entries of $\mathcal{X}^{(m)} \times \mathcal{X}^{(m)}$, respectively. To simplify notation, we drop m as a superscript when it is fixed.

For every $S, T \subseteq \mathcal{X}$, let $B(S,T)$ be the submatrix obtained from B by keeping the rows with indices in S and columns with indices in T. For every $x \in \mathcal{X}$, let $B(x,T) = B(\{x\},T)$ and $B(S,x) = B(S,\{x\})$.

The proof of Theorem 1 requires that B satisfies two properties. One is that every sufficiently large submatrix of B should have a large fraction of bad entries. This property ensures that the sum-capacity of our channel without cooperation is small (Section VI). The second property is that every submatrix of a specific type should have at least one good entry. This property enables a significantly higher sum-capacity under low-rate cooperation using the cooperation facilitator model (Section V). Lemma 2 demonstrates that these two properties can be simultaneously achieved. A proof of this and all subsequent lemmas can be found in the appendices.

Lemma 2. *Let $f, g : \mathbb{Z}^+ \to \mathbb{Z}^+$ be two functions such that $f(m) = \omega(m)$ and $g(m) = \log m + \omega(1)$. Then for every $\epsilon > 0$, there exists a sequence of $(0,1)$-matrices $\{B_m = (b_{ij})_{i,j=1}^{2^m}\}_m$ such that*

(1) for every $S, T \subseteq \mathcal{X}^{(m)}$ that satisfy $|S|, |T| \geq f(m)$,

$$\frac{|(S \times T) \cap 1_{B_m}|}{|S|\,|T|} > 1 - \epsilon;$$

that is, in every sufficiently large submatrix of B_m, the fraction of bad entries is larger than $1 - \epsilon$, and

(2) for every $x \in \mathcal{X}^{(m)}$ and $k \in \{0, 1, \ldots, 2^{m-g(m)} - 1\}$, both $B_m(x, \mathcal{X}_{m,k})$ and $B_m(\mathcal{X}_{m,k}, x)$ each contain at least one good entry, where

$$\mathcal{X}_{m,k} = \left\{ k2^{g(m)} + \ell | \ell = 1, \ldots, 2^{g(m)} \right\};$$

that is, if we break each row or column into consecutive blocks of size $2^{g(m)}$, each block contains at least one good entry.

Channel Definition: Choose functions f and g that satisfy the constraints $f(m) = \omega(m)$, $g(m) = \log m + \omega(1)$, and $\log f(m) = o(m)$. Fix a sequence of channels as defined by (4) using matrices $\{B_m\}_m$ satisfying the properties proved possible in Lemma 2 for the chosen functions f and g.

V. Inner and Outer Bounds for the CF Capacity Region

For the m^{th} channel, we show the achievability of the rate pairs $(m, m - g(m))$ and $(m - g(m), m)$, with cooperation rate $\delta_m = g(m)$. For each, we employ a blocklength-1 code ($n = 1$). Time sharing between these codes results in an inner bound for the capacity region given by

$$\begin{aligned} R_1, R_2 &\leq m, \\ R_1 + R_2 &\leq 2m - g(m). \end{aligned}$$

If $R_1 = m$, $R_2 = m - g(m)$, and $n = 1$, then the independent, uniformly distributed messages W_1 and W_2 have alphabets $\mathcal{W}_1 = \{1, \ldots, 2^m\}$ and $\mathcal{W}_2 = \{1, \ldots, 2^{m-g(m)}\}$, respectively. By the second property of our channel in Lemma 2, for every $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$, the submatrix $B_m(w_1, \mathcal{X}_{m,w_2-1})$ contains at least one good entry. Let $z = \phi(w_1, w_2)$, the output of the cooperation facilitator, be an element of $\mathcal{Z} = \{1, \ldots, 2^{g(m)}\}$ such that $(w_1, (w_2-1)2^{g(m)} + z)$ is a good entry of $B_m(w_1, \mathcal{X}_{m,w_2-1})$. If there's more than one good entry, we pick the one that results in the smallest z.

For messages $W_1 = w_1$ and $W_2 = w_2$, encoder 1 sends $x_1 = w_1$ and encoder 2 sends $x_2 = (w_2 - 1)2^{g(m)} + z$. By the definition of our channel (4), the channel output is $y = (x_1, x_2)$ with probability one, and hence zero error decoding is possible. Thus the rate pair $(m, m - g(m))$ is achievable. Note that for this achievability scheme to work, only the second encoder needs to know the value of z. A similar argument proves the achievability of $(m - g(m), m)$ and the lower bound of (2) follows.

To find an outer bound for the capacity region, we use the capacity region of the CE model [2] in a special case. Consider the situation in which encoder 1 has access to both messages and can transmit information to encoder 2 on a noiseless link of capacity δ_m. Then it is easy to see that the capacity region of this network contains the capacity region of the CF model. This situation, however, is equivalent to the CE model for $C_{12} = \delta_m$ and $C_{21} = \infty$. Hence an outer bound for the capacity region is given by the set of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(X_1; Y | X_2, U) + \delta_m, \\ R_1 + R_2 &\leq I(X_1, X_2; Y) \end{aligned}$$

for some distribution $p(u)p(x_1|u)\,p(x_2|u)$. Note that

$$\begin{aligned} I(X_1; Y | X_2, U) &\leq H(X_1) \leq m, \\ I(X_1, X_2; Y) &\leq H(X_1, X_2) \leq 2m, \end{aligned}$$

and $\delta_m = g(m)$, so the region

$$\begin{aligned} R_1 &\leq m + g(m), \\ R_1 + R_2 &\leq 2m \end{aligned}$$

is an outer bound for the CF model. Note that if we switch the roles of encoders 1 and 2, we get the outer bound

$$\begin{aligned} R_2 &\leq m + g(m), \\ R_1 + R_2 &\leq 2m. \end{aligned}$$

Since the intersection of two outer bounds is also an outer bound, the set of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1, R_2 &\leq m + g(m), \\ R_1 + R_2 &\leq 2m \end{aligned}$$

is an outer bound for the CF model as well and the upper bound of (2) follows.

VI. Inner and Outer Bounds for the IE Capacity Region

Consider the m^{th} channel of the construction in Section IV. In the case where there is no cooperation, we show that the set of all rate pairs (R_1, R_2) satisfying

$$R_1 + R_2 \leq m - g(m)$$

is an inner bound for the capacity region. To this end, we show the achievability of the rate pairs $(m - g(m), 0)$ and $(0, m - g(m))$. The achievability of all other rate pairs in the inner bound follows by time-sharing between the encoders. Similar to the achievability result of the previous section, let $n = 1$. Then $\mathcal{W}_1 = \{1, \ldots, 2^{m-g(m)}\}$ and $\mathcal{W}_2 = \{1\}$. By our channel construction, for every $w \in \mathcal{W}_1$, $B_m(\mathcal{X}_{m,w-1}, 1)$ contains at least one good entry. This means that the first column of B_m contains at least $|\mathcal{W}_1| = 2^{m-g(m)}$ good entries. Suppose encoder 1 transmits uniformly on these $2^{m-g(m)}$ good entries and encoder 2 transmits $x_2 = 1$. Then the input is always on a good entry and the channel output is the same as the channel input. Thus the pair $(m - g(m), 0)$ is achievable. A similar argument shows that the pair $(0, m - g(m))$ is achievable and the inner bound follows. We next find an outer bound for the IE capacity region.

Let Y_1 and Y_2 be the components of Y; that is, if $Y = (x_1, x_2)$, then $Y_1 = x_1$ and $Y_2 = x_2$, and if $Y = (E, E)$, then $Y_1 = Y_2 = E$. Note that $Y_1, Y_2 \in \mathcal{X} \cup \{E\}$. In the case of independent encoders, X_1 and X_2 are independent, and the distribution of Y_1 is given by

$$p(y_1) = \begin{cases} \gamma_{y_1} & y_1 \in \mathcal{X}, \\ 1 - \gamma & y_1 = E, \end{cases} \tag{5}$$

where

$$\gamma_{x_1} = p(x_1) \sum_{x_2 : b_{x_1 x_2} = 0} p(x_2),$$

for every $x_1 \in \mathcal{X}$, and $\gamma = \sum_{x_1} \gamma_{x_1}$. The capacity region for the IE model (no cooperation) is due to Ahlswede [6], [7] and Liao [8]. If $\mathscr{R}_m$ is the set of all pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(X_1; Y | X_2), \\ R_2 &\leq I(X_2; Y | X_1), \\ R_1 + R_2 &\leq I(X_1, X_2; Y) \end{aligned} \tag{6}$$

for some distribution $p(x_1)p(x_2)p(y|x_1, x_2)$ and $\mathrm{conv}(A)$ denotes the convex hull of the set A, then the capacity region is given by the closure of $\mathrm{conv}(\mathscr{R}_m)$.

If for all pairs $(R_1, R_2) \in \mathrm{conv}(\mathscr{R}_m)$, one of R_1 or R_2 is smaller than or equal to $\log 2f(m)$, then the upper bound of (3) follows, since

$$R_1 + R_2 \leq m + \log 2f(m),$$

and $\log f(m) = o(m)$. On the other hand, if there exist rate pairs $(R_1, R_2) \in \mathrm{conv}(\mathscr{R}_m)$ such that

$$R_1, R_2 > \log 2f(m), \tag{7}$$

then by (6) and (7),

$$H(X_1), H(X_2) > \log 2f(m), \tag{8}$$

and the following argument shows

$$R_1 + R_2 \leq (\sqrt{5 + 4\epsilon} - 1)m.$$

For our channel, Y, Y_1, and Y_2 are deterministic functions of (X_1, X_2), (X_1, Y_2) and (Y_1, X_2), respectively, and the bounds simplify as

$$\begin{aligned} R_1 &\leq I(X_1; Y | X_2) = H(Y_1 | X_2) \leq H(Y_1), \\ R_2 &\leq I(X_2; Y | X_1) = H(Y_2 | X_1) \leq H(Y_2). \end{aligned} \tag{9}$$

To bound $H(Y_1)$, we apply the following lemma, proved in the appendix. This lemma bounds the probability that a random variable X falls in a specific set T; the bound is given as a function of the entropy of X and the cardinality of T. For any set T, we denote its indicator function by $\mathbf{1}_T$.

Lemma 3. *Let X be a discrete random variable with distribution $p : \mathcal{X} \to \mathbb{R}_{\geq 0}$, and let T be a subset of $\mathcal{X}$. If $q : T \to \mathbb{R}_{\geq 0}$ is a function and $\alpha = \sum_{x \in T} q(x)$, then*

$$-\sum_{x \in T} q(x) \log q(x) \leq \alpha \log |T| - \alpha \log \alpha. \tag{10}$$

When $q(x) = p(x)\mathbf{1}_T(x)$, the above inequality implies

$$\alpha = \sum_{x \in T} p(x) \leq K \left(1 - \frac{H(X) - 1}{\log |\mathcal{X}|}\right), \tag{11}$$

where $K = \left(1 - \frac{\log |T|}{\log |\mathcal{X}|}\right)^{-1}$.

By (5),

$$H(Y_1) = -\sum_{x_1} \gamma_{x_1} \log \gamma_{x_1} - (1 - \gamma) \log (1 - \gamma).$$

Applying (10) from Lemma 3,

$$H(Y_1) \leq \gamma m + H(\gamma) \leq \gamma m + 1. \tag{12}$$

We next bound γ. To this end, we write each of the input distributions as a particular convex combination of uniform distributions. This is stated in the next lemma.

Lemma 4. *If X is a discrete random variable with a finite alphabet $\mathcal{X}$, then there exists a positive integer k, a sequence of positive numbers $\{\alpha_j\}_{j=1}^k$, and a collection of non-empty subsets of $\mathcal{X}$, $\{S_j\}_{j=1}^k$, such that the following properties are satisfied.*

(a) For every $j \in \{1, \ldots, k-1\}$, S_{j+1} is a proper subset of S_j.

(b) For all $x \in \mathcal{X}$,

$$p(x) = \sum_{j=1}^{k} \alpha_j \frac{\mathbf{1}_{S_j}(x)}{|S_j|}.$$

(c) For every C, $0 < C < |\mathcal{X}|$,

$$\sum_{j : |S_j| \leq C} \alpha_j \leq K \left(1 - \frac{H(X) - 1}{\log |\mathcal{X}|}\right),$$

where $K = \left(1 - \frac{\log C}{\log |\mathcal{X}|}\right)^{-1}$.

Using the previous lemma we write $p(x_1)$ and $p(x_2)$ as

$$p(x_1) = \sum_{i=1}^{k} \alpha_i^{(1)} \frac{\mathbf{1}_{S_i^{(1)}}(x_1)}{|S_i^{(1)}|},$$
$$p(x_2) = \sum_{j=1}^{l} \alpha_j^{(2)} \frac{\mathbf{1}_{S_j^{(2)}}(x_2)}{|S_j^{(2)}|}.$$

Then

$$\gamma = \sum_{x_1, x_2 : b_{x_1 x_2} = 0} p(x_1) p(x_2) = \sum_{i=1}^{k} \sum_{j=1}^{l} \alpha_i^{(1)} \alpha_j^{(2)} \beta_{ij},$$

where

$$\begin{aligned} \beta_{ij} &= \sum_{x_1, x_2 : b_{x_1 x_2} = 0} \frac{\mathbf{1}_{S_i^{(1)}}(x_1) \mathbf{1}_{S_j^{(2)}}(x_2)}{|S_i^{(1)}| |S_j^{(2)}|} \\ &= \frac{\left| \left(S_i^{(1)} \times S_j^{(2)}\right) \cap 0_{B_m} \right|}{\left|S_i^{(1)}\right| \left|S_j^{(2)}\right|} \end{aligned}$$

For every i and j, $\beta_{ij} \leq 1$. If, however, $\min\{|S_i^{(1)}|, |S_j^{(2)}|\} \geq f(m)$, then by the first property of our channel (Lemma 2),

$\beta_{ij} \le \epsilon$. Thus by part (c) of Lemma 4 and (8),

$$\begin{aligned}
\gamma &< \epsilon + \sum_{i,j:\min\{|S_i^{(1)}|,|S_j^{(2)}|\}<f(m)} \alpha_i^{(1)}\alpha_j^{(2)} \\
&= \epsilon + 1 - \sum_{i,j:\min\{|S_i^{(1)}|,|S_j^{(2)}|\}\ge f(m)} \alpha_i^{(1)}\alpha_j^{(2)} \\
&= \epsilon + 1 \\
&\quad - \left(1 - \sum_{i:|S_i^{(1)}|<f(m)} \alpha_i^{(1)}\right)\left(1 - \sum_{j:|S_j^{(2)}|<f(m)} \alpha_j^{(2)}\right) \\
&\le \epsilon + 1 - \left(1 - K_m\left(1 - \frac{H(X_1)-1}{m}\right)\right) \\
&\quad \times \left(1 - K_m\left(1 - \frac{H(X_2)-1}{m}\right)\right),
\end{aligned}$$

where $K_m = \left(1 - \frac{\log f(m)}{m}\right)^{-1}$ and $K_m \to 1$ as $m \to \infty$ since $\log f(m) = o(m)$ by assumption. Since by (6) and (7), $\log 2f(m) \le R_i \le H(X_i)$ for $i = 1, 2$,

$$\begin{aligned}
\gamma &< \epsilon + 1 - \left(1 - K_m\left(1 - \frac{R_1-1}{m}\right)\right) \\
&\quad \times \left(1 - K_m\left(1 - \frac{R_2-1}{m}\right)\right) \\
&= \epsilon + K_m\left(2 - \frac{R_1+R_2-2}{m}\right) \\
&\quad - K_m^2\left(1 - \frac{R_1-1}{m}\right)\left(1 - \frac{R_1-1}{m}\right).
\end{aligned}$$

Combining the previous inequality with (9) and (12) results in

$$\begin{aligned}
\frac{R_1}{m} &\le \epsilon + \frac{1}{m} + K_m\left(2 - \frac{R_1+R_2-2}{m}\right) \\
&\quad - K_m^2\left(1 - \frac{R_1-1}{m}\right)\left(1 - \frac{R_1-1}{m}\right) \\
&= \epsilon + \frac{1}{m} + K_m\left(2 - \frac{R_1+R_2-2}{m}\right) \\
&\quad - K_m^2\left(1 - \frac{R_1+R_2-2}{m} + \frac{(R_1-1)(R_2-1)}{m^2}\right).
\end{aligned}$$

If we let $x = \frac{R_1}{m}$ and $y = \frac{R_2}{m}$, then the previous inequality can be written as

$$\begin{aligned}
x &\le \epsilon + \frac{1}{m} + K_m\left(2 + \frac{2}{m} - x - y\right) \\
&\quad - K_m^2\left(1 + \frac{2}{m} - x - y + \left(x - \frac{1}{m}\right)\left(y - \frac{1}{m}\right)\right),
\end{aligned}$$

or

$$(x - a_m)(y + b_m) \le c_m, \tag{13}$$

where

$$\begin{aligned}
a_m &= 1 + \frac{1}{m} - \frac{1}{K_m}, \\
b_m &= -1 - \frac{1}{m} + \frac{1}{K_m} + \frac{1}{K_m^2}, \\
c_m &= -1 - \frac{2}{m} - \frac{1}{m^2} + \left(2 + \frac{2}{m}\right)\frac{1}{K_m} \\
&\quad + \left(\epsilon + \frac{1}{m}\right)\frac{1}{K_m^2} - a_m b_m.
\end{aligned}$$

By symmetry, we can also show

$$(x + b_m)(y - a_m) \le c_m. \tag{14}$$

Note that

$$\begin{aligned}
a &:= \lim_{m\to\infty} a_m = 0, \\
b &:= \lim_{m\to\infty} b_m = 1, \\
c &:= \lim_{m\to\infty} c_m = 1 + \epsilon.
\end{aligned}$$

Let S_m be the set of all nonnegative x, y that satisfy (13) and (14) and $\mathscr{S}_m$ be the set of all (mx, my) such that $(x, y) \in S_m$. Then by the arguments of this section, every $(R_1, R_2) \in \mathscr{R}_m$ that satisfies $R_1, R_2 > \log 2f(m)$ is in $\mathscr{S}_m$. As the capacity region is given by the closure of $\mathrm{conv}(\mathscr{R}_m)$, the definition of sum-capacity (1) implies

$$\begin{aligned}
\frac{1}{m}C_{\mathrm{S-IE}}^{(m)} &\le \frac{1}{m}\max_{(R_1,R_2)\in\mathrm{conv}(\mathscr{S}_m)}(R_1 + R_2) \\
&= \max_{(x,y)\in\mathrm{conv}(S_m)}(x + y).
\end{aligned}$$

Thus

$$\limsup_{m\to\infty} \frac{C_{\mathrm{S-IE}}^{(m)}}{m} \le \lim_{m\to\infty}\max_{(x,y)\in\mathrm{conv}(S_m)}(x + y). \tag{15}$$

To find the limit on the right hand side, we make use of the following lemma proved in the appendix.

Lemma 5. *Suppose $\{a_m\}_{m=1}^\infty$, $\{b_m\}_{m=1}^\infty$ and $\{c_m\}_{m=1}^\infty$ are three sequences of real numbers such that $\lim_{m\to\infty} a_m = a$, $\lim_{m\to\infty} b_m = b$, $\lim_{m\to\infty} c_m = c$, where*

$$b, c, a + b, ab + c > 0,$$

and

$$\sqrt{(a+b)^2 + 4c} > b + \frac{c}{b}.$$

For every positive integer m, let S_m be defined as above. Then

$$\lim_{m\to\infty}\max_{(x,y)\in\mathrm{conv}(S_m)}(x + y) = a - b + \sqrt{(a+b)^2 + 4c}.$$

It is easy to see that the sequences above satisfy the assumptions of Lemma 5. Thus

$$\limsup_{m\to\infty}\frac{C_{\mathrm{S-IE}}^{(m)}}{m} \le \sqrt{5 + 4\epsilon} - 1,$$

Therefore, for all but finitely many m,

$$C_{\mathrm{S-IE}}^{(m)} \le (\sqrt{5 + 4\epsilon} - 1)m.$$

VII. Conclusion

In this paper, we present a new model for cooperation and study its benefits in the case of the encoders of a DM-MAC. Specifically, we present channels for which the gain in sum-capacity is "almost" exponential in the cooperation rate. The CF model can be generalized to other network settings, and its study is subject to future work.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant Numbers CCF-1321129, CCF-1018741, CCF-1038578, and CNS-0905615.

Appendix A
Proof of Lemma 2

We use the probabilistic method [15]. We assign a probability to every $2^m \times 2^m$ $(0,1)$-matrix and show that the probability of a matrix having both properties is positive for sufficiently large m; hence, there exists at least one such matrix. Fix $\epsilon > 0$, and let $B_m = (b_{ij})_{i,j=1}^{2^m}$ be a random matrix with $b_{ij} \overset{\text{iid}}{\sim}$ Bernoulli (p), where $1-\epsilon < p < 1$. Let

$$\Gamma_m = \left\{ S : S \subseteq \mathcal{X}^{(m)}, |S| \geq f(m) \right\}$$

For every $S, T \in \Gamma_m$, define the event

$$E_{S,T}^{(m)} = \left\{ \frac{|(S \times T) \cap 1_{B_m}|}{|S|\,|T|} \leq 1-\epsilon \right\}.$$

It follows

$$\begin{aligned}
&\Pr\left(\bigcup_{S,T\in\Gamma} E_{S,T}^{(m)} \right) \\
&\quad \leq \sum_{S,T\in\Gamma} \Pr\left(E_{S,T}^{(m)} \right) \\
&\quad = \sum_{S,T\in\Gamma} \Pr\left(|(S\times T) \cap 1_{B_m}| \leq (1-\epsilon)\,|S||T| \right) \\
&\quad = \sum_{S,T\in\Gamma} \sum_{k=0}^{\lfloor (1-\epsilon)|S||T| \rfloor} \binom{|S||T|}{k} p^k (1-p)^{|S||T|-k} \\
&\quad = \sum_{i,j=f(m)}^{2^m} \binom{2^m}{i}\binom{2^m}{j} \sum_{k=0}^{\lfloor (1-\epsilon)ij \rfloor} \binom{ij}{k} p^k (1-p)^{ij-k}.
\end{aligned}$$

Suppose $\{X_l\}_{l=1}^L$ is a set of independent random variables such that for each l, $X_l \in [a_l, b_l]$ with probability one. If $S = \sum_{l=1}^L X_i$, Hoeffding's inequality [16] states that for every u smaller or equal to $\mathbb{E}S$, we have

$$\Pr(S \leq u) \leq e^{-\frac{2(\mathbb{E}S-u)^2}{\sum_{l=1}^L (b_l-a_l)^2}}.$$

If $\{X_l\}_{l=1}^{ij}$ is a set of ij independent Bernoulli(p) random variables, then for every l, $0 \leq X_l \leq 1$, and

$$(1-\epsilon)ij \leq pij = \mathbb{E}\left[\sum_{l=1}^{ij} X_l \right].$$

Thus Hoeffding's inequality implies

$$\begin{aligned}
\sum_{k=0}^{\lfloor (1-\epsilon)ij \rfloor} \binom{ij}{k} p^k (1-p)^{ij-k} &= \Pr\left(\sum_{l=1}^{ij} X_l \leq (1-\epsilon)ij \right) \\
&\leq e^{-2(p-1+\epsilon)^2 ij}.
\end{aligned}$$

Since $\binom{2^m}{i} \leq 2^{mi}$,

$$\begin{aligned}
&\sum_{i,j=f(m)}^{2^m} \binom{2^m}{i}\binom{2^m}{j} \sum_{k=0}^{\lfloor (1-\epsilon)ij \rfloor} \binom{ij}{k} p^k (1-p)^{ij-k} \\
&\quad \leq \sum_{i,j=f(m)}^{2^m} 2^{m(i+j)} e^{-2(p-1+\epsilon)^2 ij} \\
&\quad = \sum_{i,j=f(m)}^{2^m} e^{(i+j)m\ln 2 - 2(p-1+\epsilon)^2 ij}.
\end{aligned}$$

If we define $h : \mathbb{Z}^2 \to \mathbb{R}$ as

$$h(i,j) = (i+j)m\ln 2 - 2(p-1+\epsilon)^2 ij,$$

then for $j \geq f(m)$,

$$\begin{aligned}
&h(i+1,j) - h(i,j) \\
&\quad = m\ln 2 - 2(p-1+\epsilon)^2 j \\
&\quad \leq m\ln 2 - 2(p-1+\epsilon)^2 f(m) \\
&\quad = f(m)\left(\frac{m}{f(m)} \ln 2 - 2(p-1+\epsilon)^2 \right).
\end{aligned}$$

By assumption,

$$\lim_{m\to\infty} \frac{m}{f(m)} = 0,$$

so there exists M_1 such that for all $m > M_1$,

$$\frac{m}{f(m)} < \frac{2}{\ln 2}(p-1+\epsilon)^2.$$

Therefore, for $m > M_1$ and $y \geq f(m)$, h is decreasing with respect to i. As h is symmetric with respect to i and j, for $m > M_1$ and $i \geq f(m)$, we also have $h(i,j+1) - h(i,j) < 0$. Thus h is a decreasing function in i and j for $m > M_1$ and $i, j \geq f(m)$. Hence for $m > M_1$,

$$\begin{aligned}
&\sum_{i,j=f(m)}^{2^m} e^{(i+j)m\ln 2 - 2(p-1+\epsilon)^2 ij} \\
&\quad \leq (2^m - f(m) + 1)^2 e^{2mf(m)\ln 2 - 2(p-1+\epsilon)^2 (f(m))^2} \\
&\quad < e^{2m(1+f(m))\ln 2 - 2(p-1+\epsilon)^2 (f(m))^2} \\
&\quad = e^{2(f(m))^2 \left(\left(1+\frac{1}{f(m)}\right)\frac{m}{f(m)}\ln 2 - 2(p-1+\epsilon)^2 \right)}.
\end{aligned}$$

The exponent of the right hand side of the previous inequality,

$$2\,(f(m))^2 \left(\left(1 + \frac{1}{f(m)}\right) \frac{m}{f(m)} \ln 2 - 2(p-1+\epsilon)^2 \right),$$

goes to $-\infty$ as m approaches infinity, so

$$\lim_{m\to\infty} \Pr\left(\bigcup_{S,T\in\Gamma} E_{S,T}^{(m)} \right) = 0.$$

This means that the probability that the fraction of bad entries in a sufficiently large submatrix is less than $1-\epsilon$ is going to zero.

Next, we calculate the probability that B_m doesn't satisfy the second property. For every $x \in \mathcal{X}^{(m)}$ and $k \in \{1, \ldots, 2^{m-g(m)}\}$, define the event

$$E_{x,k}^{(m)} = \left\{ \left(0_{B_m(x,\mathcal{X}_{m,k})} \cup 0_{B_m(\mathcal{X}_{m,k},x)}\right) \cap 0_{B_m} = \emptyset \right\}.$$

The probability that for every x and k the sets $B_m(x, \mathcal{X}_{m,k})$ and $B_m(\mathcal{X}_{m,k}, x)$ don't have at least one good entry equals

$$\begin{aligned}
&\Pr\left(\bigcup_{x,k} E_{x,k}^{(m)}\right) \\
&\quad \le \sum_{x \in \mathcal{X}^{(m)}} \sum_{k=1}^{2^{m-g(m)}} \Pr\left(0_{B_m(x,\mathcal{X}_{m,k})} \cap 0_{B_m} = \emptyset\right) \\
&\quad + \sum_{x \in \mathcal{X}^{(m)}} \sum_{k=1}^{2^{m-g(m)}} \Pr\left(0_{B_m(\mathcal{X}_{m,k},x)} \cap 0_{B_m} = \emptyset\right) \\
&\quad = 2^{2m-g(m)+1} p^{2^{g(m)}} \\
&\quad = 2^{2^{g(m)}\left(\frac{2m-g(m)+1}{2^{g(m)}} + \log p\right)}.
\end{aligned}$$

Since $m = o(2^{g(m)})$, the exponent of the right hand side of the previous inequality,

$$2^{g(m)}\left(\frac{2m-g(m)+1}{2^{g(m)}} + \log p\right),$$

goes to $-\infty$ as $m \to \infty$. This implies that

$$\Pr\left(\bigcup_{x,k} E_{x,k}^{(m)}\right)$$

goes to zero as $m \to \infty$. Thus, by the union bound the probability that the matrix doesn't satisfy either of these properties is going to zero. Therefore, for large enough m, almost every $(0,1)$-matrix satisfies both properties, though we only need one such matrix.

Appendix B
Proof of Lemma 3

For the first part, if $\alpha = 0$, then $q(x) = 0$ for every $x \in T$ and both sides equal zero. Otherwise,

$$\begin{aligned}
-\sum_{x \in T} q(x) \log q(x) &= -\alpha \sum_{x \in T} \frac{q(x)}{\alpha} \log \frac{q(x)}{\alpha} - \alpha \log \alpha \\
&\le \alpha \log |T| - \alpha \log \alpha,
\end{aligned}$$

since $q(x)/\alpha$ is a probability mass function with entropy $\sum_{x \in T} \frac{q(x)}{\alpha} \log \frac{\alpha}{q(x)}$ and alphabet size $|T|$.

For the second part, if

$$q(x) = p(x) \mathbf{1}_T(x),$$

then by the previous inequality,

$$\begin{aligned}
-\sum_{x \in T} p(x) \log p(x) &= -\sum_{x \in T} q(x) \log q(x) \\
&\le \alpha \log |T| - \alpha \log \alpha,
\end{aligned}$$

where

$$\alpha = \sum_{x \in T} q(x) = \Pr(x \in T).$$

Similarly, replacing $\mathcal{X} \setminus T$ with T results in

$$\begin{aligned}
&-\sum_{x \in \mathcal{X} \setminus T} p(x) \log p(x) \\
&\qquad \le (1-\alpha) \log |\mathcal{X} \setminus T| - (1-\alpha) \log (1-\alpha).
\end{aligned}$$

Adding the previous two inequalities gives

$$\begin{aligned}
H(X) &\le \alpha \log |T| + (1-\alpha) \log |\mathcal{X} \setminus T| + H(\alpha) \\
&\le \alpha \log |T| + (1-\alpha) \log |\mathcal{X}| + 1.
\end{aligned}$$

Therefore,

$$\frac{H(X)}{\log |\mathcal{X}|} \le 1 + \frac{1}{\log |\mathcal{X}|} - \left(1 - \frac{\log |T|}{\log |\mathcal{X}|}\right) \alpha,$$

and

$$\alpha \le \frac{1 - \frac{H(X)-1}{\log |\mathcal{X}|}}{1 - \frac{\log |T|}{\log |\mathcal{X}|}}.$$

Appendix C
Proof of Lemma 4

Let k be the cardinality of the range of $p : \mathcal{X} \to \mathbb{R}$. Then there exists a sequence $\{x_j\}_{j=1}^k$ such that

$$\{p(x) | x \in \mathcal{X}\} = \{p(x_j) | 1 \le j \le k\},$$

and

$$0 < p(x_1) < \cdots < p(x_k) \le 1.$$

For j, $1 \le j \le k$, define

$$S_j = \{x \in \mathcal{X} | p(x) \ge p(x_j)\},$$

and let $S_{k+1} = \emptyset$. Then for j, $1 \le j \le k$, $S_{j+1} \subseteq S_j$ (part a) and

$$S_j \setminus S_{j+1} = \{x \in \mathcal{X} | p(x) = p(x_j)\}.$$

Thus, the number of $x \in \mathcal{X}$ such that $p(x) = p(x_j)$ equals $|S_j \setminus S_{j+1}|$. For $j \in \{2, \ldots, k\}$, define

$$\alpha_j = |S_j| \left(p(x_j) - p(x_{j-1})\right),$$

and let $\alpha_1 = |S_1| p(x_1)$.

In part (b), the left hand side simplifies as

$$\begin{aligned}
\sum_{j=1}^k \alpha_j \frac{\mathbf{1}_{S_j}(x)}{|S_j|} &= \sum_{j=1}^k \left(p(x_j) - p(x_{j-1})\right) \mathbf{1}_{S_j}(x) \\
&= \sum_{j=1}^k p(x_j) \mathbf{1}_{S_j \setminus S_{j+1}}(x) \\
&= p(x).
\end{aligned}$$

In (c), if the set $\{j | 1 \le j \le k, |S_j| \le C\}$ is empty, then there's nothing to prove. Otherwise, it's a nonempty subset of

$\{1, \ldots, k\}$ and thus has a minimum, which we call j^*. Then

$$\begin{aligned}\sum_{j:|S_j|\le C}\alpha_j &= \sum_{j=j^*}^{k}\alpha_j \\ &= \sum_{j=j^*}^{k}|S_j|\left(p(x_j)-p(x_{j-1})\right) \\ &= \sum_{j=j^*}^{k}|S_j\setminus S_{j+1}|p(x_j) - |S_{j^*}|p(x_{j*-1}) \\ &= \sum_{x\in S_{j^*}} p(x) - |S_{j^*}|p(x_{j*-1}) \\ &\le \sum_{x\in S_{j^*}} p(x).\end{aligned}$$

By (11) of Lemma 3,

$$\begin{aligned}\sum_{x\in S_{j^*}} p(x) &\le \frac{1}{1-\frac{\log|S_{j^*}|}{\log|\mathcal{X}|}}\left(1-\frac{H(X)-1}{\log|\mathcal{X}|}\right) \\ &\le \frac{1}{1-\frac{\log C}{\log|\mathcal{X}|}}\left(1-\frac{H(X)-1}{\log|\mathcal{X}|}\right),\end{aligned}$$

since $|S_{j^*}| \le C$.

Appendix D
Proof of Lemma 5

Prior to proving Lemma 5, we state and prove the following lemma.

Lemma 6. *Suppose a, b, and c are three real numbers such that b, c, $a+b$, and $ab+c$ are positive, and*

$$\sqrt{(a+b)^2+4c} > b + \frac{c}{b}.$$

Let S be the set of all pairs (x,y) such that $x, y \ge 0$, and

$$\begin{cases}(x-a)(y+b) & \le c, \\ (x+b)(y-a) & \le c.\end{cases}$$

If x_0 is the unique positive solution to the equation

$$(x-a)(x+b) = c,$$

then

$$\max_{(x,y)\in\mathrm{conv}(S)} (x+y) = 2x_0.$$

Proof: Since

$$(x-a)(y+b) - (x+b)(y-a) = (a+b)(x-y)$$

and $a+b>0$, the set S can be written as $S = S_1 \cup S_2$ (Figure 3), where S_1 is the set of all pairs (x,y) such that $0 \le x \le y$ and

$$(x+b)(y-a) \le c,$$

and S_2 is the set of all pairs (x,y) such that $0 \le y \le x$ and

$$(x-a)(y+b) \le c.$$

The intersection of S_1 and S_2 consists of all pairs (x,x) such that $0 \le x \le x_0$ where

$$x_0 = \frac{a-b+\sqrt{(a+b)^2+4c}}{2}.$$

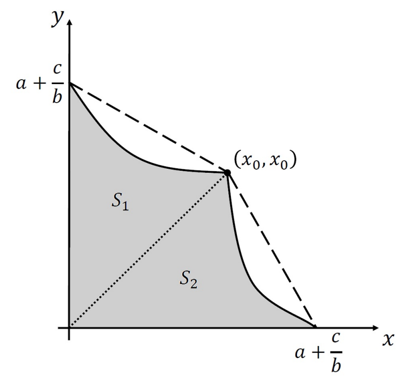

Figure 3. The sets S_1 and S_2 (gray area), and their convex hulls.

Note that since b, c and $ab+c$ are positive,

$$\sqrt{(a+b)^2+4c} < a+b+\frac{2c}{b},$$

so $0 < x_0 < a + \frac{c}{b}$. The convex hull of S_1 consists of all pairs (x,y) such that $0 \le x \le y$ and

$$\left(a+\frac{c}{b}-x_0\right)x + x_0 y \le \left(a+\frac{c}{b}\right)x_0,$$

and the convex hull of S_2 consists of all pairs (x,y) such that $0 \le y \le x$ and

$$x_0 x + \left(a+\frac{c}{b}-x_0\right)y \le \left(a+\frac{c}{b}\right)x_0.$$

Note that $\mathrm{conv}(S_1) \cup \mathrm{conv}(S_2)$ is the region bounded by the axes $y=0$, $x=0$, and the piecewise linear function

$$h(x) = \begin{cases}\frac{x_0-a-\frac{c}{b}}{x_0}x + a + \frac{c}{b} & 0 \le x \le x_0, \\ \frac{x_0}{x_0-a-\frac{c}{b}}x - \frac{\left(a+\frac{c}{b}\right)x_0}{x_0-a-\frac{c}{b}} & x_0 < x \le a+\frac{c}{b}.\end{cases}$$

Since $2x_0 \ge a + \frac{c}{b}$ by assumption,

$$\frac{x_0-a-\frac{c}{b}}{x_0} \ge \frac{x_0}{x_0-a-\frac{c}{b}}.$$

This means the slope of h is decreasing, or equivalently, h is a concave function. Thus $\mathrm{conv}(S_1) \cup \mathrm{conv}(S_2)$ is convex. But

$$S \subseteq \mathrm{conv}(S_1) \cup \mathrm{conv}(S_2) \subseteq \mathrm{conv}(S),$$

so

$$\mathrm{conv}(S) = \mathrm{conv}(S_1) \cup \mathrm{conv}(S_2).$$

This implies

$$\max_{(x,y)\in\mathrm{conv}(S)} (x+y) = 2x_0.$$

□

Using this lemma, we may prove Lemma 5. There exists a positive M such that for every $m \ge M$,

$$b_m, c_m, a_m+b_m, a_m b_m + c_m > 0$$

and

$$\sqrt{(a_m+b_m)^2+4c_m} - b_m - \frac{c_m}{b_m} > 0.$$

Let $x_0^{(m)}$ and x_0 be the unique positive solutions to the equations

$$(x_0^{(m)} - a_m)(x_0^{(m)} + b_m) = c_m$$

and

$$(x_0 - a)(x_0 + b) = c.$$

Since $x_0^{(m)}$ and x_0 are continuous functions of (a_m, b_m, c_m) and (a, b, c), respectively, we have

$$\lim_{m\to\infty} x_0^{(m)} = x_0.$$

Thus by Lemma 6,

$$\begin{aligned}\lim_{m\to\infty} \max_{(x,y)\in\mathrm{conv}\left(S^{(m)}\right)} (x+y) &= \lim_{m\to\infty} 2x_0^{(m)} \\ &= 2x_0 \\ &= a - b + \sqrt{(a+b)^2 + 4c}.\end{aligned}$$

REFERENCES

[1] G. Kramer, I. Marić, and R. D. Yates, "Cooperative communications," *Foundations and Trends in Networking*®, vol. 1, no. 3-4, pp. 271–425, 2006.

[2] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 441–445, 1983.

[3] R. Dabora and S. D. Servetto, "Broadcast channels with cooperating decoders," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5438–5454, 2006.

[4] I. Marić, R. D. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3536–3548, 2007.

[5] H. H. Permuter, S. Shamai, and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6379–6396, 2011.

[6] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory*, 1971.

[7] ——, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, pp. 805–814, 1974.

[8] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elect. Eng., Univ. Hawaii, 1972.

[9] T. Ho, M. Effros, and S. Jalali, "On equivalence between network topologies," in *Proc. Allerton Conf. Communication, Control, and Computing*, 2010.

[10] S. Jalali, M. Effros, and T. Ho, "On the impact of a single edge on the network coding capacity," in *Information Theory and Applications Workshop*, 2011.

[11] M. Langberg and M. Effros, "Network coding: Is zero error always possible?" in *Proc. Allerton Conf. Communication, Control, and Computing*, 2011.

[12] ——, "Source coding for dependent sources," in *Proc. IEEE Information Theory Workshop*, 2012.

[13] E. J. Lee, M. Langberg, and M. Effros, "Outer bounds and a functional study of the edge removal problem," in *Proc. IEEE Information Theory Workshop*, 2013.

[14] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York: Cambridge University Press, 2012.

[15] J. Spencer, *Ten Lectures on the Probabilistic Method*. Philadelphia: SIAM, 1994.

[16] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.